AFRL-IF-RS-TR-2002-237
Final Technical Report
September 2002

ROUGH 'N' READY: A MEETING RECORDER AND BROWSER

BBN Technologies

Sponsored by
Defense Advanced Research Projects Agency
DARPA Order No. F301

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2002-237 has been reviewed and is approved for publication

APPROVED: Mark D. Foresti

MARK D. FORESTI
Project Engineer

FOR THE DIRECTOR: Michael Talbert

MICHAEL L. TALBERT, Technical Advisor
Information Technology Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
SEPTEMBER 2002

3. REPORT TYPE AND DATES COVERED
Final Jun 97 - Jun 00

4. TITLE AND SUBTITLE
ROUGH 'N' READY: A MEETING RECORDER AND BROWSER

5. FUNDING NUMBERS
C - F30602-97-C-0253
PE - 62301E
PR - F301
TA - 01
WU - 00

6. AUTHOR(S)
John Makhoul and Francis Kubala

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
BBN Technologies
10 Moulton Street
Cambridge Massachusetts 02138

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington Virginia 22203-1714

Air Force Research Laboratory/IFTD
525 Brooks Road
Rome New York 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-IF-RS-TR-2002-237

11. SUPPLEMENTARY NOTES

Air Force Research Laboratory Project Engineer: Mark D. Foresti/IFTD/(315) 330-2233/ Mark.Foresti@rl.af.mil

12a. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The objective of this effort is to integrate and enhance existing technologies in speech recognition, speaker identification, and topic classification to provide cost-effective transcription, structural summarization, and retrieval of user-specified aspects of meetings. A software system consisting of a meeting recorder and browser was designed and developed to provide a higher level view of collaborative meetings, co-locational or distributed and a way to browse through and listen to those parts which are most relevant to the user.

14. SUBJECT TERMS
Browser, Meeting Recorder, Collaboration, Speech Processing, Speech Integration

15. NUMBER OF PAGES
24

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Table of Contents

List of Figures

List of Tables

1 Executive Summary

This is the final report for Contract F30602-97-C-0253, DARPA Order No. F301, entitled "Rough'n'Ready: A Meeting Recorder and Browser", for the period from 30 June 1997 to 30 September 2000. Note that the original expiration date for this contract was 29 June 2000. A no-cost extension to the contract, through 30 September 2000, was granted in order to support BBN attendance at the DARPA-sponsored Information Management Principle Investigators' meeting in September 2000.

At the outset of this contract, we set four major goals:

- **Integration** of several diverse speech and language processing components to extract a *rich-content* representation of audio data
- **Presentation** of the rich-content meta-data to a browser client over the Internet
- **Demonstration** of sustained real-time indexing of continuous audio input
- **Transfer** of Rough'n'Ready technology to real-world applications for the U.S. Government

Over the original 36-month period of this contract, we made substantial and demonstrable progress against each of these four objectives.

1.1 Component Integration

Eight advanced speech and language technologies were successfully taken from the research environment and developed into runtime components that work together to produce a rich-content representation of audio. By the end of the contract, we had successfully integrated the following components into a unified system:

- Speaker Change Detection
- Large Vocabulary Speech Recognition
- Speaker Clustering
- Speaker Identification
- Named-Entity Extraction
- Story Segmentation
- Topic Classification
- Probabilistic Information Retrieval

Integration of such diverse speech and language technologies is a novel approach for extracting information from speech and audio sources. Rough'n'Ready has clearly demonstrated that there is considerable benefit for using all available acoustic and linguistic extraction technology at once. Individually, each technology is imperfect, but together, they offer a rich and redundant representation of speech that is useful for many important applications.

1.2 Browser Presentation

The rich-content meta-data extracted by Rough'n'Ready is stored in a relational database and is made available to users using a common browser over the Internet. In this way, users are able to navigate and search within large archives of audio and video data with the same relative ease in which large text collections on the Web are searched today.

Rough'n'Ready is set up as a Web Application Service, which means that the end user needs nothing more than a browser (Internet Explorer 5+) to interact with the service. There is no client software to download or register beforehand. We employ XML and XSLT as middleware intermediaries between the database and the client browser. These highly structured and *non-proprietary protocols* allow Rough'n'Ready to present its metadata to the client over the Internet with rich display features and complex interactions using only ubiquitous HTTP protocols. This is a modern Web-oriented architecture that effectively leverages many existing and future Internet standards.

1.3 Real-time Demonstration

When this contract began in 1997, we were running the BBN Byblos recognizer at a throughput rate of 10 times real-time (10xRT => ten times slower than the real-time duration of the input audio). Note that the Byblos large vocabulary speech recognizer is by far the most compute-intensive component within Rough'n'Ready. By the end of the contract, we had achieved 0.8xRT throughput on a COTS dual-processor Pentium III PC and we had decreased the absolute Word Error Rate (WER). It is a significant accomplishment to achieve a >10x speedup while improving recognition accuracy at the same time.

Furthermore, at the beginning of the contract, the Byblos recognizer processed its input in a batch mode; meaning that it produced its automatic transcription only after it had completed the processing for an entire audio session. Thus, even with a Byblos recognizer capable of real-time throughput, an application or a user would have to wait until the audio session was complete before any of the meta-data could be examined. For example, a user would need to wait 30 minutes to see the transcript of a 30-minute program. By the end of this contract, however, we had reengineered the Byblos recognition engine to run on continuous input while outputting its results incrementally. Thus, a user could begin viewing the partial transcript of a session within a minute of its capture.

These are breakthrough achievements that put Rough'n'Ready technology in a position to be deployed in field trials for some interesting transcription and audio monitoring applications.

1.4 Technology Transfer

By the end of this contract, we had made three significant technology transfers into the government and commercial arenas.

BBN is under contract with the Foreign Broadcast Information Service (FBIS) to build and deploy several Rough'n'Ready systems to support manual selection, transcription, and translation of broadcast news in several languages around the world. For this application and government customer, Rough'n'Ready technology has been customized and renamed, ***OASIS***, which stands for *Open Audio Source Information System.* BBN built a prototype system, to be deployed in Bangkok, Thailand, that monitors accented English from Radio and TV channels in

Pakistan and India. The automatic OASIS transcript is used to help the human monitors more efficiently make selections of important stories from audio sources. Secondly, the transcript is used to speed up the process of manual transcription of these selected stories before they are fed into the FBIS product stream. In the future, we are under contract to develop similar capabilities for South African accented-English and to support manual translation of audio in Spanish, Arabic, and Chinese. This application for FBIS is a direct transition of advanced technology from the DARPA IM program into a fielded operational application for the U.S. Government. It is perhaps, the most reliable and visible indicator of the success of the Rough'n'Ready contract.

In the commercial sector, BBN licensed its Rough'n'Ready technology to L&H for continuing development of the component technologies into products and applications. This agreement is designed to drive Rough'n'Ready capabilities more rapidly into marketplace through strategic partnering.

OASIS audio indexing capabilities have begun transitioning to the new DARPA TIDES program; to be part of a larger loosely-integrated system called, TIDES Portal. The initial target for TIDES Portal was to participate in an Integrated Feasibility Experiment (IFE) called, RIMPAC-2000, which is a U.S. Navy-sponsored, multi-national, joint services exercise that was conducted on the island of Hawaii in June 2000. Our part of the exercise focused on gathering information in the field from open sources to help in humanitarian assistance operations. The setting for the IFE was Marine camp in a harsh desert-like environment on the northwest side of the island. OASIS and the other components in the TIDES Portal system were all loaded onto portable servers and laptops and brought to the camp for the exercise. The system was designed to cache information from the Internet locally so that it could continue to function without requiring a continuous Internet connection. This exercise was an important as exposure for R&D staff to real-world operational conditions.

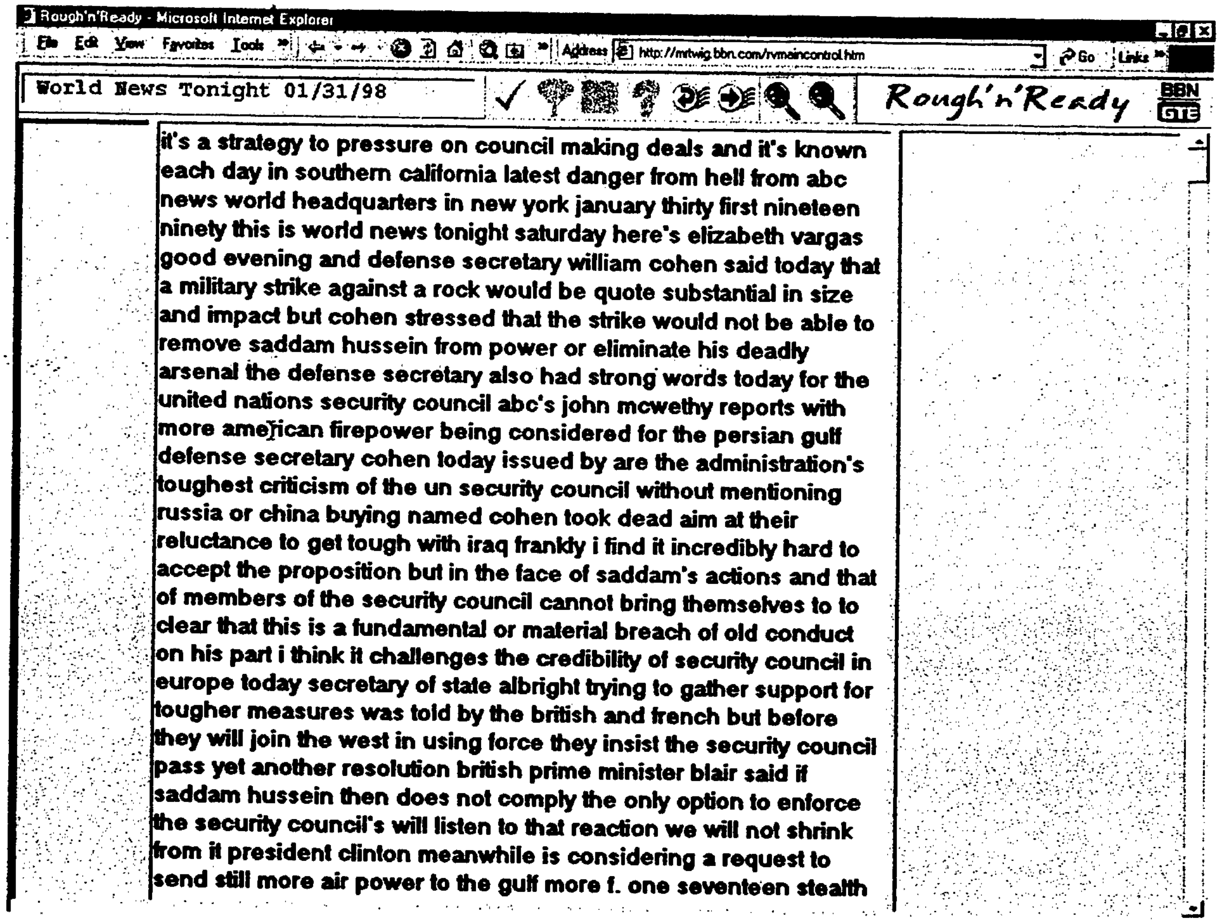
2 Component Integration for Rich-Content Extraction

The primary goal of this contract at the outset was to integrate five speech and language processing components into a single system that would be capable of providing a content-based index into spoken language. The five components to be integrated were, speech recognition, speaker identification, name spotting, topic classification, and information retrieval. BBN has developed advanced capabilities in each of these areas with significant U.S. government support over many years. Integration of such a diverse collection of speech and language technologies had never been attempted before.

The motivating vision for the integration was the conjecture that these components, working together, could automatically create a rich index into the content of spoken language that would be useful for managing the information contained in large collections of digital audio recordings, including those from radio, television, and telephone sources. Without such a rich-content index, spoken language is very difficult to review or retrieve for content since there is no way to search or navigate within raw audio/video other than to listen to the material with a real-time playback device.

If only an automatic speech recognition (ASR) component is available to index the spoken language audio track, simple keyword searches can then be made, but retrieval and navigation of content from ASR alone is not adequate. This is because ASR output alone is not organized enough for the human reviewer to efficiently browse, retrieve, select, summarize, and organize the content of the spoken material.

In Figure 1 below, we show the raw output of a state-of-the-art ASR system (Byblos) on a broadcast news program. This automatic transcript contains about 20% word errors. It is fairly difficult to read and retain mentally because it has no sentence or paragraph boundaries, and no case or punctuation. More importantly, it has no boundaries between the various news stories so that they cannot be indexed for retrieval the way textual documents are.


Rough'n'Ready - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Address
http://mtwig.bbn.com/rvmaincontrol.htm
Go
Links
Rough'n'Ready
BBN
GTE

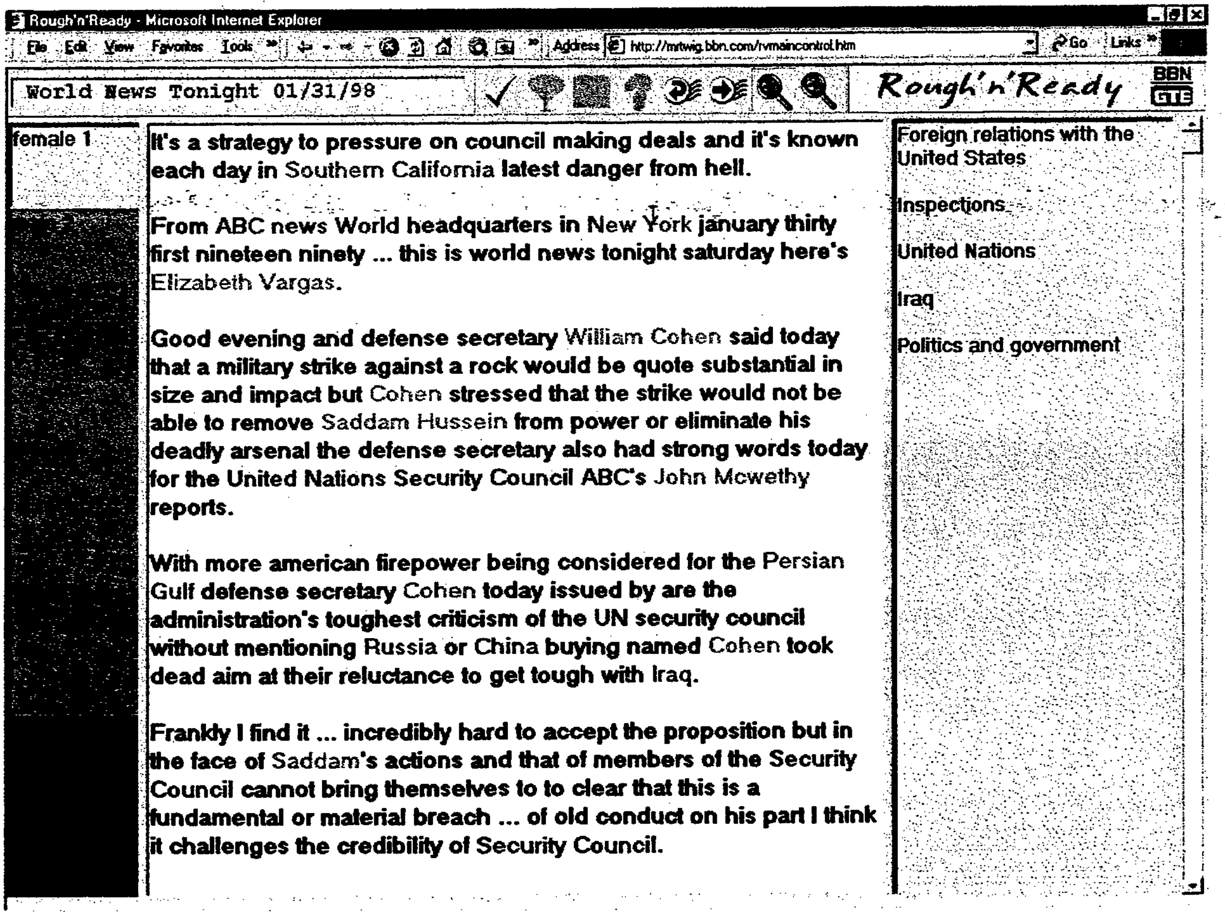
World News Tonight 01/31/98

it's a strategy to pressure on council making deals and it's known each day in southern california latest danger from hell from abc news world headquarters in new york january thirty first nineteen ninety this is world news tonight saturday here's elizabeth vargas good evening and defense secretary william cohen said today that a military strike against a rock would be quote substantial in size and impact but cohen stressed that the strike would not be able to remove saddam hussein from power or eliminate his deadly arsenal the defense secretary also had strong words today for the united nations security council abc's john mcwethy reports with more american firepower being considered for the persian gulf defense secretary cohen today issued by are the administration's toughest criticism of the un security council without mentioning russia or china buying named cohen took dead aim at their reluctance to get tough with iraq frankly i find it incredibly hard to accept the proposition but in the face of saddam's actions and that of members of the security council cannot bring themselves to to clear that this is a fundamental or material breach of old conduct on his part i think it challenges the credibility of security council in europe today secretary of state albright trying to gather support for tougher measures was told by the british and french but before they will join the west in using force they insist the security council pass yet another resolution british prime minister blair said if saddam hussein then does not comply the only option to enforce the security council's will listen to that reaction we will not shrink from it president clinton meanwhile is considering a request to send still more air power to the gulf more f. one seventeen stealth

Figure 1. The raw output of an automatic speech recognizer is difficult to read and organize mentally.

Even though the precise location of keywords in an audio track can be made with the aid of a speech recognizer, the human reviewer still has a considerable amount of cognitive work to do to understand the topical content of the surrounding speech in detail and to find boundaries for the passage of interest so that it can be extracted, or marked for some other use. The lack of boundaries, case and punctuation, as well as the presence of transcription errors all conspire to impede understanding of the underlying speech content. The aim of Rough'n'Ready is to take this *rough* transcription and make it *ready* for browsing by enriching it with a variety of content-based features that can be automatically extracted from the audio track alone.

In Figure 2 below, we show the rich-content output from the integrated technologies applied to the same passage shown in Figure 1. On the left column in Figure 2, we see that the boundaries between speakers have created paragraph-like blocks of text in the transcription. In addition, the speakers are identified either by name or simply by gender, and passages spoken by the same speaker are clustered together. Within the text itself, the proper names of people, places, and organizations are highlighted with color and capitalization. Together, these additional rich-content features transform the raw transcription into a much more usable form. With the help provided by these additional features, the reader is less burdened by the effort of deciphering the automatic transcript and is thereby more able to concentrate on the task of finding specific content in the speech.


Rough'n'Ready - Microsoft Internet Explorer
File Edit View Favorites Tools
Address http://mrtwig.bbn.com/rvmaincontrol.htm
Go
Links
World News Tonight 01/31/98
Rough'n'Ready
BBN
GTE
female 1

It's a strategy to pressure on council making deals and it's known each day in Southern California latest danger from hell.

From ABC news World headquarters in New York january thirty first nineteen ninety ... this is world news tonight saturday here's Elizabeth Vargas.

Good evening and defense secretary William Cohen said today that a military strike against a rock would be quote substantial in size and impact but Cohen stressed that the strike would not be able to remove Saddam Hussein from power or eliminate his deadly arsenal the defense secretary also had strong words today for the United Nations Security Council ABC's John Mcwethy reports.

With more american firepower being considered for the Persian Gulf defense secretary Cohen today issued by are the administration's toughest criticism of the UN security council without mentioning Russia or China buying named Cohen took dead aim at their reluctance to get tough with Iraq.

Frankly I find it ... incredibly hard to accept the proposition but in the face of Saddam's actions and that of members of the Security Council cannot bring themselves to to clear that this is a fundamental or material breach ... of old conduct on his part I think it challenges the credibility of Security Council.

Foreign relations with the United States

Inspections

United Nations

Iraq

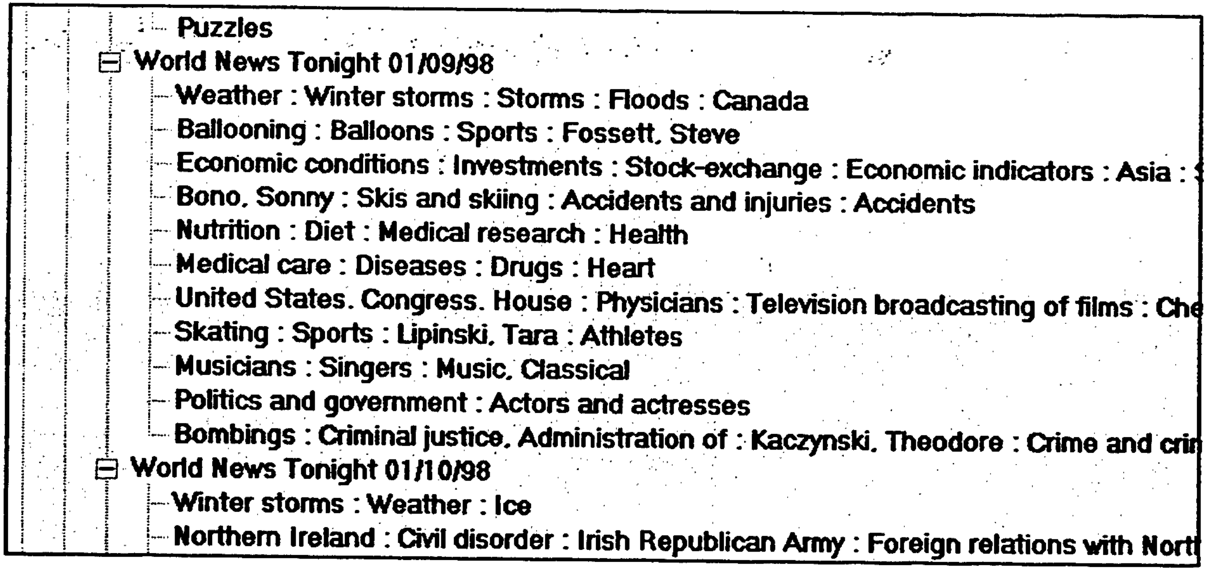
Politics and government

Figure 2. The *rich-content* features produced by the integrated technologies of the Rough'n'Ready system help the human reader organize the information in speech and to locate specific content.

On the right-hand column of Figure 2 is shown a set of topic labels that the system has automatically selected to summarize the content of the news story shown in the center panel. These predefined topic labels are selected by the system from over five thousand possible labels known to the system. The system selects all topics from this large set that are judged to be relevant to the words in the story. Just as importantly, the system rejects all of the other topic labels that are judged *not* relevant to the story. Topic classification in this manner is unique to BBN as far as we know. No other work that we are aware of employs such a large set of topics and attempts to label stories with a variable number of them.

Topic classification of speech is a very important content-based feature for several reasons. It permits the end-user to look for content by an abstract concept rather than simply by keywords alone. Note that the actual terms in the topic labels are not required to occur in the text of the transcript. In this way, topic classification generalizes beyond any form of keyword-based search.

The topic classifier is also used in Rough'n'Ready to segment the continuous speech input into thematic passages or stories. This is a breakthrough capability that transforms continuous speech into *Spoken Documents*, which enables effective Information Retrieval to be performed on audio data. Lacking segmentation of stories, an IR engine would need to break the audio transcriptions into arbitrary blocks against which to match the query. These arbitrary blocks of text would not be thematically coherent and therefore Information Retrieval performance would suffer.



- Puzzles
- World News Tonight 01/09/98
  - Weather : Winter storms : Storms : Floods : Canada
  - Ballooning : Balloons : Sports : Fossett. Steve
  - Economic conditions : Investments : Stock-exchange : Economic indicators : Asia :
  - Bono. Sonny : Skis and skiing : Accidents and injuries : Accidents
  - Nutrition : Diet : Medical research : Health
  - Medical care : Diseases : Drugs : Heart
  - United States. Congress. House : Physicians : Television broadcasting of films : Che
  - Skating : Sports : Lipinski. Tara : Athletes
  - Musicians : Singers : Music. Classical
  - Politics and government : Actors and actresses
  - Bombings : Criminal justice. Administration of : Kaczynski. Theodore : Crime and crin
- World News Tonight 01/10/98
  - Winter storms : Weather : Ice
  - Northern Ireland : Civil disorder : Irish Republican Army : Foreign relations with Nort

Figure 3. ***Spoken Documents*** **are automatically created from continuous speech input by story segmentation and topic classification.**

The topic labels also provide a concise and informative summary of the content in speech as is illustrated in Figure 3 above. Each line below the node labeled, *World News Tonight 01/09/98,* represents a distinct story that was reported in the evening broadcast of ABC's World News Tonight program on January 9, 1998. These descriptive summary lines are formed from the collection of topic labels automatically chosen by Rough'n'Ready to classify the topics in the story. They are the same labels shown in the right column of Figure 2, which is a view of the Rough'n'Ready transcription of a single story. From Figure 3, it is readily apparent that with the story labels, a quick 30-second reading of the story labels provides a reasonable gist of 30 minutes of news. This is a major enabling capability for users who must deal with audio or video data for its spoken content.

The automatic segmentation and labeling of continuous speech by the Rough'n'Ready Indexer is a major technical achievement. This work grew out of a Master's Thesis by Amit Srivastava at Northeastern University. The Thesis, entitled *"Story Segmentation in Audio Indexing"* is included as an attachment to this report.

3 Web-based Audio Browser

An important goal of the Rough'n'Ready project was to expose its capabilities over the Internet to allow and encourage its use in collaboration with other information technology research. As reported previously, we abandoned our original architectural design for the system when we discovered that it did not scale easily to the Web. The initial architecture was built upon a collection of Microsoft's ActiveX components, many of which, we created. Our new approach leverages the rapidly evolving XML family of inter-process middleware services that are being developed as an open standard that is free of commercial control (but open to commercial influence).

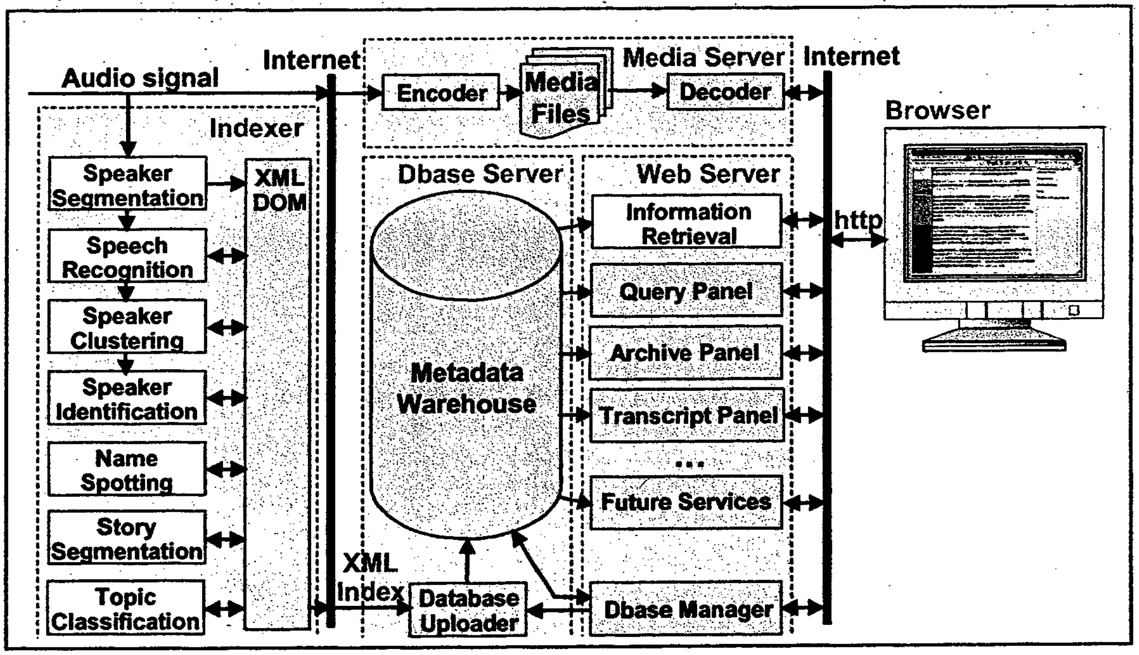


Figure 4. Architecture of the Rough'n'Ready audio indexing system. The system leverages HTTP and XML to simplify inter-process communications over the Internet. Components depicted as yellow rectangles denote speech and language technologies developed at BBN with many years of DARPA support.

The current architecture of Rough'n'Ready is strongly database-centric, as shown in Figure 4 above. The function of the indexer (shown on the left in Figure 2) is to automatically extract as many content-based features as is possible from spoken language and to pass them to the database as structured metadata. In our current architecture, the metadata is passed as XML, which many commercial database services are beginning to accept as native communications. This greatly simplifies communications between indexing services like Rough'n'Ready and the data warehouses that they populate.

In the middle section of Figure 4, we illustrate a collection of three discrete services that comprise the Rough'n'Ready server in our current architecture. The Media Server handles the source audio data, compressing it as it arrives from the source and streaming it to the client/user on demand. We have used both RealPlayer7 and Microsoft's NetShow Services for our Media Server.

The Dbase Server is the heart of the system. It contains all of the content-based metadata that has been automatically extracted by the indexer. Each extracted feature is indexed by time back to the source media so that user-initiated playback of specific content is possible. For our current architecture, any commercial RDBMS is suitable, but we have done all of our development work using Microsoft's SQL Server.

The third component of the server complex is the Web server, which is responsible for mediating interactions between the user and the database. COTS Web servers are becoming very reliable and scalable, and the details of the Web server used are not important for the Rough'n'Ready architecture. We are currently using Microsoft's ISAPI as the common denominator for our Web server. We are using MS ASP (Active Server Pages) and ADO (Active Data Objects) technologies with considerable success. The ease with which we are able to prototype client User Interfaces and link them to the metadata is very promising. We do not spend a lot of effort building system infrastructure, leaving more of our effort to concentrate on the value-added activities of research and algorithm development. Even though the server-side technologies that we use now are proprietary, they do not impact the browser client, which means that any fully capable Web server available today could serve the same role adequately.

The browser client itself is another COTS component and is not strictly part of the Rough'n'Ready system. The design principle operative here is that the Rough'n'Ready index should be accessible by any up-to-date commercial or open-source browser that supports XML and XSLT. To that end, we have avoided any requirement that the client must download any registered components or accept cookies from the Rough'n'Ready server. Access to the content-indexed Rough'n'Ready database is initiated simply by pointing a browser to the Web server's URL.

The complex visualizations of the data and the wide variety of interactions between the user and the database are all made possible by sending XML/XSLT data between the client and server under control of the ASPs residing on the server. Since XML is an extensible protocol and because it is human-readable as well as machine-readable, the job of defining communications interfaces between the client and server is made much easier than for previous distributed computing strategies. This kind of IPC is platform and programming language independent. It also supports flexible APIs that can be modified additively without requiring that all clients be updated at the same time. Finally, since the XML family of middleware technologies is governed by an independent technical body (the World Wide Web Consortium or W3C) and is being driven forward by the immediate needs of e-commerce, we can be reasonably assured that this enabling technology will get better and better over time, and that it will not fall into the grips of proprietary interests that restrict its growth.

4 Real-Time Processing

By the end of this contract, we had completed the real-time acoustic pipeline for Rough'n'Ready. This allows the system to process audio input continuously from radio or TV without falling behind. This is an important milestone for the contract – and one that will bring collateral advantages for every future research effort or tech-transfer opportunity. Real-time processing is a natural requirement for any application that deals with audio or other time-based media.

The Rough'n'Ready acoustic pipeline now consists of a cascade of exactly six programs in series that communicate via TCP/IP sockets. The 6 stages function as follows:

1. Speech/non-speech detection
2. Speaker change detection
3. Acoustic fast match
4. N-best generation
5. N-best rescoring
6. Speaker identification

This sequence of acoustic processing programs pass data and control parameters through sockets to their neighbors in the pipeline. This Inter-Process Communications (IPC) strategy allows the programs to run on different processors without any change to the components. We have exploited this capability by using a dual-processor PC to achieve real-time throughput without sacrificing recognition accuracy. Since the socket protocols are location-independent, the individual stages of the pipeline could be run on several machines connected by a network just as easily.

In Figure 5 below, we show our progress in real-time processing over the course of the last two years of the Rough'n'Ready contract. When we began this contract in July 1997, we started with a state-of-the-art version of our large vocabulary speech recognition engine that ran in about 100 times real-time (100xRT). This was the version of Byblos that we used for the annual DARPA-sponsored *Hub-4* technology evaluations based on the broadcast news domain. As shown below, we began formal tracking of our real-time speed and accuracy in September of 1998. At that time, we benchmarked our system at 7.5xRT on a 450 MHz Pentium II PC.

Since 1998, we have increased the throughout of the Rough'n'Ready BYBLOS recognizer by a factor of 9. At the same time, the Word Error Rate (WER) has actually decreased by about 14% relative. This chart shows clearly that we have achieved our real-time goals without sacrificing anything in the way of accuracy. We have actually improved our recognition accuracy over the same time that we achieved a 9-fold reduction in runtime. That is a rare achievement in large vocabulary transcription work.

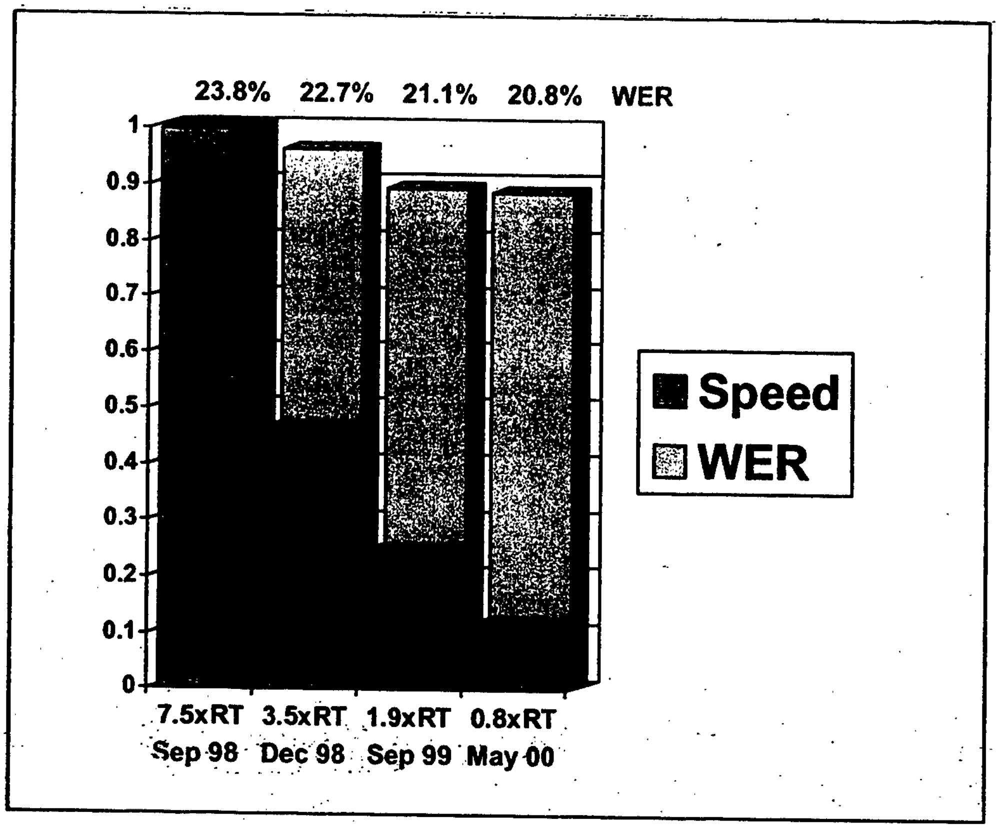


Figure 5. Throughput speed and recognition accuracy benchmarks for the Rough'n'Ready BYBLOS ASR engine measured on the 1997 Hub-4 test set over the last two years. Speed and accuracy numbers are normalized to 1.0 for the September 1998 benchmark.

These benchmarks were achieved with a speaker-independent acoustic model, a 65K word dictionary, and a language model containing about 30M ngrams. The final benchmark in May 2000, was achieved on a dual processor 866 MHz Pentium III PC. The remarkable speedups were achieved primarily through algorithmic improvements to the ASR engine. Roughly 3 of the 9 factors of speedup were due to improvements in hardware over the period of the contract.

We measure our real-time ASR performance as the total elapsed time needed to transcribe an audio episode divided by the duration of the episode. We begin timing at the moment the first input segment is available to the system and we stop the clock when the last of the input has been completely transcribed. Our real-time measurements are always made on a single high-end COTS PC. Note that this manner of measuring real-time includes whatever processing latency the system needs to process the input continuously. It also includes all initialization time, speaker change detection and utterance segmentation time. This is a realistic and conservative approach to measuring real-time performance.

In table 1 below, we show accuracy (WER) and throughput (xRT) numbers for all Hub-4 evaluation test sets from 1996 through 1999. Note that the throughput numbers are correlated to accuracy but throughput varies far less than WER. This is a satisfying and useful result.

Hub-4 Test Set	WER	xRT
1996	28.8	0.86
1997	20.8	0.81
1998-1	21.2	0.79
1998-2	18.1	0.74
1999-1	24.1	0.79
1999-2	21.6	0.78

Table 1. Summary of accuracy (WER) and throughput speed (xRT) for each of the Hub-4 evaluation test sets from 1996 to 1999.

At the same time that we were improving the throughput of our recognizer, we were also improving the acoustic indexer in other important ways. At the beginning of the Rough'n'Ready contract, the BYBLOS recognizer was composed of over 35 individual stages that were distributed over a set of compute servers connected via Ethernet. Today, it runs on a single machine, taking digital audio as input and producing an XML index as output. In addition, the original segmentation and recognition processes were non-causal, requiring many passes over the audio data and each stage of the process was required to complete its pass over the data before any subsequent process could begin. Today, the 6 stages of the acoustic pipeline are causal with a short delay of less than a minute, which permits the whole indexing engine to keep up with continuous speech input producing an index in an incremental fashion. Taken together, these improvements have made Rough'n'Ready more operationally believable and have prepared it for deployment as a prototype production system in the FBIS worldwide audio transcription workflow.

5 Technology Transfer

5.1 Foreign Broadcast Information Service

The Rough'n'Ready technology suite has attracted a real-world customer from the US government sector. We are under contract with FBIS (Foreign Broadcast Information Service) to deliver systems to be deployed in Thailand, Panama, and South Africa. The Bangkok and South African systems are intended for use on accented-English radio and TV broadcasts from India and Pakistan, and South Africa. The system slated for Panama will be in trained to index Spanish broadcasts. These early prototype systems will operate 24x7 and will be fully automated. The initial function of the systems is to make the selection of important news stories from TV and radio more efficient for FBIS production personnel. Workflow studies have already been conducted at the Bangkok facility and were judged successful. We anticipate staff efficiency improvements on the order of factors of 2-3 over the current manual audio review process.

The Rough'n'Ready contract contributed to this important tech-transfer opportunity in several crucial ways. Most important among its contributions are the basic information management capabilities themselves provided by the many DARPA-sponsored BBN technologies that are integrated into the Rough'n'Ready system. But real-time processing capability and the Web-capable architecture are also vital contributions made by this contract to the transfer of DARPA-sponsored research to a real-world customer.

Longer-term plans for the FBIS deployment call for new language capabilities, including Mandarin Chinese and Arabic, in the year 2001. Data collection efforts for these languages are underway. Taken together, the technology transfer effort that is currently underway to FBIS represent a substantial success for the DARPA IC&V program that supported the initial Rough'n'Ready demonstration and for all the preceding DARPA programs that contributed to the many component technologies in Rough'n'Ready today.

5.2 DARPA TIDES Portal - RIMPAC 2000

From its inception, one of the major goals of the Rough'n'Ready contract was to set up a collaborative demonstration of its information management capabilities on the Web that included other DARPA funded research. In June 2000, we achieved this goal by making a convincing demonstration of Rough'n'Ready capabilities over the Web at the Strong Angel exercise in Hawaii in collaboration with MITRE, Lincoln Laboratory, and Global InfoTek. This exercise was conceived and supported by the DARPA TIDES (Translingual Information Detection, Extraction, and Summarization) program. The collaboration produced a Web-based TIDES Portal giving humanitarian assistance personnel (both civilian and military) access to timely worldwide news with advanced content-based information management tools. Rough'n'Ready supplied the Portal with the capability to index audio sources (radio and TV) by its content.

For this exercise, a Rough'n'Ready system, located at BBN in Cambridge, Massachusetts, automatically indexed incoming worldwide news texts and radio broadcasts on a daily basis. Users in Hawaii could browse the indexed database directly over the Web. In addition, a local database in Hawaii could be updated daily by replication of the new entries in the database located in Cambridge. This strategy was designed to permit the uninterrupted operation of the service even in the event of a network failure at the base camp in Hawaii.

This exercise pushed the Rough'n'Ready tech-base toward greater operational capability in two important directions. First, we inserted text documents into the indexer for name extraction and topic classification before uploading to the database. For the first time, the Rough'n'Ready system allowed users to browse and retrieve information from a database of content features extracted from mixed audio and text sources. The user had no need to know the original medium of the source in order to get useful retrievals from the indexed corpus. This was a breakthrough capability that will be leveraged for all mixed-media language technology research going forward.

Another important new influence that the TIDES Portal effort had upon the Rough'n'Ready effort was its emphasis upon fresh daily updates of processed news sources. As a consequence, we spent some effort getting the indexer and server systems to function in a 24x7 operational manner, capturing the audio and text data on an schedule and updating the metadata warehouse in a completely automatic procedure. This gave us a good deal of early practical experience in designing systems for the operational environment. Our current Web-based demo is continually updated in this manner, processing audio sources once a day and text source every 30 minutes on average. We have also scaled up our database schema so that we can handle 10's of thousands of text and audio documents easily. Current work underway (under the TIDES Portal contract) will allow us to scale the system up to millions of documents.

Finally, the TIDES Portal exercise gave us some practical experience in connecting diverse information management systems that were developed under a variety of different contexts. In particular, we learned how to easily export our automatically extracted features to another automated service for subsequent value-added processing. For the case in point, we exported an XML index containing only our extraction of named entities (a simple reduction of our normal Rough'n'Ready index) to a geo-spatial display system (GeoNODE) developed at MITRE.

These two systems were developed independently over several years at two different sites, with no knowledge of each other's APIs. Nonetheless, it was trivial for us to define a simple XML message that contained all of the information extracted by Rough'n'Ready needed for display on the GeoNODE maps. Since both systems had their own Web servers in place, this new communication between them required only a small modification to the communications services on both sides. More importantly, these modifications were general in nature and made no use of knowledge that was specific to Rough'n'Ready or GeoNODE. After these modifications were made, Rough'n'Ready became capable of sending named-entity information to any downstream service and GeoNODE became capable of receiving it from any upstream service that offered it. This was a powerful demonstration of the collaborative possibilities that lie ahead for confederations of advanced information management services exposed to each other via the Internet.

List of Publications

The following publications are included as attachments to this report.

[1] John Makhoul, Francis Kubala, Tim Leek, Daben Liu, Long Nguyen, Richard Schwartz, and Amit Srivastava, "***Speech and language Technologies for Audio Indexing and Retrieval,***" Proceedings of the IEEE, Vol. 88, No. 8, August 2000.

[2] Francis Kubala, Sean Colbath, Daben Liu, Amit Srivastava, and John Makhoul, "***Integrated Technologies for Indexing Spoken Language,***" Communications of the ACM, Vol. 43, No. 2, February 2000.

[3] Sean Colbath, Francis Kubala, Daben Liu, and Amit Srivastava, "***Spoken Documents: Creating Searchable Archives from Continuous Audio,***" 33rd Hawaii International Conference on System Sciences (HICSS), January 2000.

[4] Daben Liu and Francis Kubala, "***Fast Speaker Change Detection for Broadcast News Transcription and Indexing,***" Eurospeech 99, Budapest, Hungary, September 1999.

[5] Amit Srivastava, "***Story Segmentation in Audio Indexing,***" Master of Science Thesis, Department of Electrical and Computer Engineering, Northeastern University, Boston Massachusetts, August 1999

[6] Sean Colbath and Francis Kubala, "***Rough'n'Ready: A Meeting Recorder and Browser,***" 2nd Annual Workshop on Perceptual User Interfaces, San Francisco, California, November 1998.